

Analysis of Count Data by Transmuted Geometric Distribution

Subrata Chakraborty¹, Deepesh Bhati^{2,*}

¹Department of Statistics, Dibrugarh University, Assam, India

²Department of Statistics, Central University of Rajasthan, Rajasthan, India

ARTICLE INFO

Article History

Received 30 Nov 2017

Accepted 14 Sep 2018

Keywords

Transmuted geometric distribution
 EM algorithm
 Likelihood Ratio test
 Rao score's test
 Wald's test

ABSTRACT

Transmuted geometric distribution (*TGD*) was recently introduced and investigated by Chakraborty and Bhati [Stat. Oper. Res. Trans. 40 (2016), 153–176]. This is a flexible extension of geometric distribution having an additional parameter that determines its zero inflation as well as the tail length. In the present article we further study this distribution for some of its reliability, stochastic ordering and parameter estimation properties. In parameter estimation among others we discuss an EM algorithm and the performance of estimators is evaluated through extensive simulation. For assessing the statistical significance of additional parameter (α), Likelihood ratio test, the Rao's score tests and the Wald's test are developed and its empirical power via simulation are compared. We have demonstrate two applications of (*TGD*) in modeling real life count data.

© 2019 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

Chakraborty and Bhati [1] recently introduced the transmuted geometric distribution $TGD(q, \alpha)$ using the quadratic rank transmutation techniques of Shaw and Buckley [2]. It may be noted that though there is a large number of new continuous distribution in statistical literature which are derived using the rank transmutation technique but $TGD(q, \alpha)$ is the first discrete distribution derived using this technique. Chakraborty and Bhati [1] investigated various distributional properties, showed applicability of $TGD(q, \alpha)$ in modeling aggregate loss, claim frequency data from automobile insurance and demonstrated the feasibility of $TGD(q, \alpha)$ as count regression model by considering data from health sector. In the current article, we discussed some additional theoretical and applied aspects of $TGD(q, \alpha)$, which are structured as follows. In Section 3, we present various reliability properties and stochastic ordering of $TGD(q, \alpha)$. In Section 4, comparative study of maximum likelihood estimator (MLE) obtained numerically and through EM Algorithm are presented through simulation, whereas in Section 5, detailed hypothesis testing is discussed considering three Wald's, Rao's Score and Likelihood Ratio test (LRT) for testing $\alpha = 0$. To illustrate the applicability of (*TGD*) models in different disciplines other than those discussed in Chakraborty and Bhati [1], we consider two real data sets and compare them with different family of distributions in Section 6. Finally, some conclusions and comments are presented in Section 7.

2. TRANSMUTED GEOMETRIC DISTRIBUTION ($TGD(q, \alpha)$)

A random variable (rv) X is said to follow **Transmuted geometric distribution** (*TGD*) with two parameters q and α , in short, $TGD(q, \alpha)$ if its probability mass function (pmf) is given by

$$p_y = (1 - \alpha)q^y(1 - q) + \alpha(1 - q^2)q^{2y}, \quad y = 0, 1, \dots \quad (1)$$

The corresponding survival function (sf) is written as

$$\bar{F}_Y(y) = (1 - \alpha)q^y + \alpha q^{2y}, \quad y = 0, 1, \dots, \quad (2)$$

where $0 < q < 1$, $-1 < \alpha < 1$. Following distributional characteristics are presented in Chakraborty and Bhati [1].

*Corresponding author. Email: deepesh.bhati@curaj.ac.in

1. For $\alpha = 0$, (1) reduces to $GD(q)$ with pmf $p_y = (1 - q)q^y$, $y = 0, 1, \dots$, $0 < q < 1$.
2. For $\alpha = -1$, (1) reduces to a special case of the **Exponentiated Geometric distribution (EGD)** of Chakraborty and Gupta (2015) with power parameter equal to 2. This is the **distribution of the maximum** of two independent and identical distributed (iid) $GD(q)$ rvs.
3. For $\alpha = 1$, (1) reduces to $GD(q^2)$ with pmf $(1 - q^2)q^{2y}$, which is the **distribution of the minimum** of two iid $GD(q)$ rvs.
4. For $0 < \alpha < 1$ ($-1 < \alpha < 0$) the $TGD(q, \alpha)$ distribution with pmf given in (1), the ratio p_y/p_{y-1} , $y = 1, 2, \dots$, forms a monotone increasing (decreasing) sequence.
5. $TGD(q, \alpha)$ is unimodal with a nonzero mode for $-1 < \alpha < -(q(2 + q))^{-1}$ provided $q > 0.414$.
6. The probability generating function (PGF) of $TGD(q, \alpha)$ is given by

$$G_Y(z) = \frac{(1 - q)(1 - \alpha q(1 - z) - q^2 z)}{(1 - qz)(1 - q^2 z)}, \quad |q^2 z| < 1.$$

7. The r^{th} factorial moment of $Y \sim TGD(q, \alpha)$ is given by

$$\mathbb{E}(Y_{(r)}) = (1 - \alpha)r! \left(\frac{q}{1 - q}\right)^r + \alpha r! \left(\frac{q^2}{1 - q^2}\right)^r,$$

where $Y_{(r)} = Y(Y - 1) \dots (Y - r + 1)$.

3. RELIABILITY PROPERTIES AND STOCHASTIC ORDERING

There are several situations in reliability where continuous time is not a good scale to measure the lifetime, in production we may be interested in how many units are produced by the machine before failure or health insurance companies are interested how long a patient stays in hospital before discharge/death. In such situations, the discrete hazard rate functions can be used to model ageing properties of discrete random lifetimes. The different hazard rate functions of TGD model and associated results are as follows:

3.1. Reliability Properties

3.1.1. Hazard rate function and its classification

The hazard rate function $r_X(x)$ for $X \sim TGD(q, \alpha)$ is given as

$$\begin{aligned} r_X(x) &= \frac{P(X = x)}{S_X(x)} = \frac{(1 - \alpha)q^x(1 - q) + \alpha(1 - q^2)q^{2x}}{(1 - \alpha)q^x + \alpha q^{2x}}, \\ &= \frac{(1 - \alpha)(1 - q) + \alpha q^x(1 - q^2)}{(1 - \alpha) + \alpha q^x}. \end{aligned}$$

The hazard rate function of $TGD(q, \alpha)$ is plotted in Figure 1 for various values of parameters to investigate the monotonic properties and it is clear that the hazard rate of $TGD(q, \alpha)$ is **increasing** for $-1 < \alpha < 0$, **decreasing** when $0 < \alpha < 1$ and **constant** if $\alpha = 0$ or 1. Also it can be seen that even when $\alpha \neq 1$, the hazard rate approaches to a constant as y increases. The smaller the value of q the faster is the rate of stabilization of the hazard rate.

Theorem 1. The $TGD(q, \alpha)$ has increasing, decreasing and constant hazard rate for $-1 < \alpha < 0$, $0 < \alpha < 1$ and $\alpha = 0$ or 1 respectively.

Proof. The hazard rate function of $TGD(q, \alpha)$ is given as

$$\begin{aligned} r_Y(y) &= \frac{(1 - \alpha)(1 - q) + \alpha q^y(1 - q^2)}{(1 - \alpha) + \alpha q^y} \\ &= 1 - q \frac{(1 - \alpha) + \alpha q^{y+1}}{(1 - \alpha) + \alpha q^y}. \end{aligned}$$

But $q \frac{(1 - \alpha) + \alpha q^{y+1}}{(1 - \alpha) + \alpha q^y}$ is a decreasing (increasing) function of y for $-1 < \alpha < 0$ ($0 < \alpha < 1$). Hence $r_Y(y)$ is increasing (decreasing) function of y for $-1 < \alpha < 0$ ($0 < \alpha < 1$). Constant hazard rates are obtained as $r_Y(y) = 1 - q$ for $\alpha = 0$ and $r_Y(y) = 1 - q^2$ for $\alpha = 1$. \square

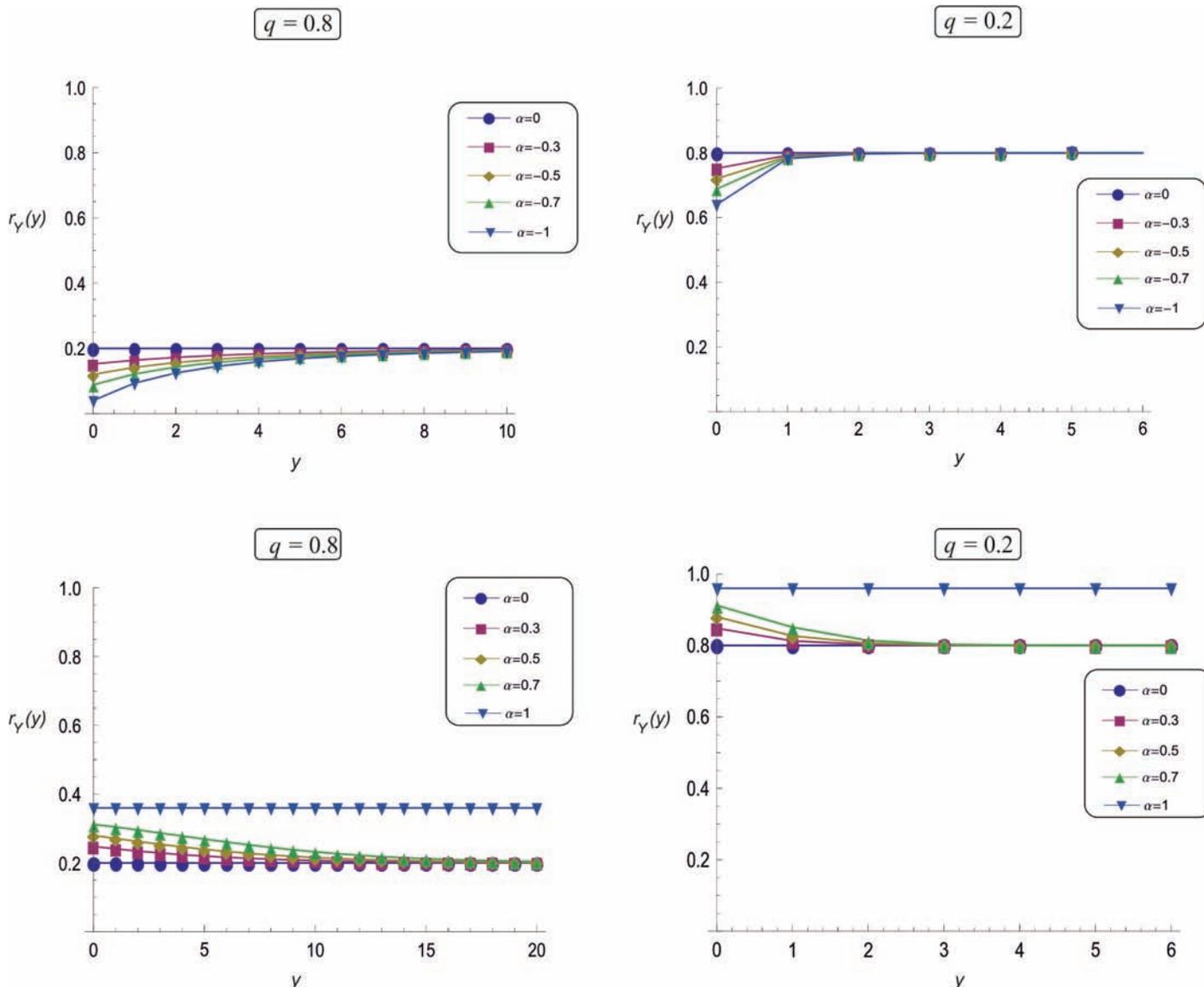


Figure 1 | Hazard rate function plots of $TGD(q, \alpha)$.

Remark. The hazard rate of $TGD(q, \alpha)$ clearly obeys $r_Y(y) \leq 1 - q$ for $-1 \leq \alpha \leq 0$ and $1 - q \leq r_Y(y) \leq 1 - q^2$ for $0 \leq \alpha \leq 1$.

3.1.2. Mean residual life

Kemp (2004) presented various characterization of discrete lifetime distribution among them the mean residual life (MRL) or life expectancy is an important characteristic, for TGD , the closed expression for MRL is given as

$$L_Y(y) = \mathbb{E}(Y - y | Y \geq y) = \frac{1}{S_Y(y)} \sum_{j>y} S_Y(j) = \frac{q((1+q)(1-\alpha) + \alpha q^{y+1})}{(1-q^2)(1-\alpha + \alpha q^y)}. \tag{3}$$

Theorem 2. The MRL function given in (3) is monotone decreasing (increasing) function of y depending on $-1 < \alpha < 0$ ($0 < \alpha < 1$).

Proof. It can be easily be seen that

$$\Delta L_Y(y) = L_Y(y+1) - L_Y(y) = \frac{(1-\alpha)\alpha q^{y+1}}{(1+q)(1-\alpha(1-q^y))(1-\alpha(1-q^{y+1}))}.$$

For any choice of $\alpha \in (-1, 1)$ and $q \in (0, 1)$, the denominator terms $(1 - \alpha(1 - q^y))$ and $(1 - \alpha(1 - q^{y+1}))$ are always positive. Moreover, since $q \in (0, 1)$, therefore $\Delta L_Y(y) < 0$ for $-1 < \alpha < 0$ indicates decreasing MRL, whereas $\Delta L_Y(y) > 0$ for $0 < \alpha < 1$ indicates increasing MRL. \square

3.2. Stochastic Ordering

Many times there is a need of comparing the behavior of one rv with the other. Shaked and Shanthikumar [3] has given many comparisons such as likelihood ratio order (\leq_{lr}), the stochastic order (\leq_{st}), the hazard rate order (\leq_{hr}), the reversed hazard rate order (\leq_{rh}) and the expectation order (\leq_E) having various applications in different context.

Theorem 3. Let Y be a rv following $TGD(q, \alpha)$ and X be geometric rv with parameter p . Then $R(z) = P(Y = z)/P(X = z)$ is an increasing(decreasing) function of z for $-1 < \alpha < 0$ ($0 < \alpha < 1$) respectively that is $X \leq_{lr} Y$ ($X \geq_{lr} Y$).

Proof. Since $R(z) = 1 + \alpha((1 + q)q^z - 1)$. Thus, we have $R(z) \leq (\geq) R(z + 1)$ for $-1 < \alpha < 0$ ($0 < \alpha < 1$) for any $q \in (0, 1)$. \square

Corollary. Following results are direct implications of Theorem 3.

1. $X \leq_{st} (\geq_{st}) Y$ that is, $P(X \geq z) \leq (\geq) P(Y \geq z)$ for $-1 < \alpha < 0$ ($0 < \alpha < 1$) respectively and for all z .
2. $Y \leq_{hr} (\geq_{hr}) X$ that is, $P(Y = z)/P(Y \geq z) \leq (\geq) P(X = z)/P(X \geq z)$ for $-1 < \alpha < 0$ ($0 < \alpha < 1$) respectively and for all z .
3. $X \leq_{rh} (\geq_{rh}) Y$ that is, $P(X = z)/P(X \leq z) \leq (\geq) P(Y = z)/P(Y \leq z)$ for $-1 < \alpha < 0$ ($0 < \alpha < 1$) respectively and for all z .
4. $X \leq_E (\geq_E) Y$ that is, $E(X) \leq (\geq) E(Y)$ for $-1 < \alpha < 0$ ($0 < \alpha < 1$) respectively and for all z .

Theorem 4. Let Y_1 and Y_2 be $TGD(q_1, \alpha)$ and $TGD(q_2, \alpha)$ respectively. Then $Y_2 \leq_{st} Y_1$ iff $q_1 \leq q_2$ for $0 < \alpha < 1$.

Proof. We know that $Y_2 \leq_{st} Y_1$ if $P(Y_2 \geq y) \leq P(Y_1 \geq y)$ for all y , hence for $TGD(q, \alpha)$ with $P(Y \geq y) = (1 - \alpha)q^{2y} + \alpha q^y$ and it is clearly seen that for $0 < \alpha < 1$

$$(1 - \alpha)q_1^{2y} + \alpha q_1^y \leq (1 - \alpha)q_2^{2y} + \alpha q_2^y \quad \forall y \quad \text{iff} \quad q_1 \leq q_2.$$

Hence $Y_2 \leq_{st} Y_1$. \square

4. PARAMETER ESTIMATION AND THEIR COMPARATIVE EVALUATION

Estimates of the parameters q and α of $TGD(q, \alpha)$ model can be computed by following five methods (i) sample proportion of 1s and 0s method, (ii) sample quantiles, (iii) method of moments and finally (iv) ML method and (v) ML via EM Algorithm. Moreover, in this section we carry out comparative study of ML estimator obtained numerically and via EM Algorithm utilizing initial estimate from one of the first three methods.

4.1. From Sample Proportion of 1's and 0's

If p_0, p_1 be the known observed proportion of 0's and 1's in the sample, then the parameters q and α can be estimated by solving the equations

$$p_0 = (1 - \alpha)(1 - q) + \alpha(1 - q^2) \quad \text{and} \quad p_1 = (1 - \alpha)q(1 - q) + \alpha q^2(1 - q^2).$$

4.2. From Sample Quantiles

If t_1, t_2 be two observed points such that $F_Y(t_1) = \gamma_1, F_Y(t_2) = \gamma_2$, then the two parameters q and α can be estimated by solving the simultaneous equations

$$\gamma_1 = 1 + (\alpha - 1)q^{t_1+1} - \alpha q^{2(t_1+1)} \quad \text{and} \quad \gamma_2 = 1 + (\alpha - 1)q^{t_2+1} - \alpha q^{2(t_2+1)}.$$

4.3. Methods of Moments

Denoting the first and second observed raw moments by m_1 and m_2 respectively, the moment estimates can be obtained by

1. Either solving the following two equations simultaneously

$$\frac{q(1-\alpha)+q^2}{1-q^2} = m_1 \quad \text{and} \quad \frac{q\left((1+q)^3 - \alpha(q(3q+2)+1)\right)}{(1-q^2)^2} = m_2,$$

2. or by the minimization method proposed by Khan *et al.* [4] by minimizing $(\mathbb{E}(Y) - m_1)^2 + (\mathbb{E}(Y^2) - m_2)^2$ with respect to q and α

$$\left(\frac{q(1-\alpha)+q^2}{1-q^2} - m_1\right)^2 + \left(\frac{q\left((1+q)^3 - \alpha(q(3q+2)+1)\right)}{(1-q^2)^2} - m_2\right)^2.$$

4.4. ML Method

Let $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ be a sample of n observations drawn from TGD distribution, and $\Theta = (q, \alpha)^T$ be the parametric vector. The log-likelihood (LL) function for the corresponding sample is

$$l = \log L = n \log(1-q) + \log(q) \sum_{i=1}^n y_i + \sum_{i=1}^n \log((1-\alpha) + \alpha q^{y_i}(1+q)), \tag{4}$$

and the score function $U(\Theta, \mathbf{y}) = \left(\frac{\partial l}{\partial q}, \frac{\partial l}{\partial \alpha}\right)^T$ can be obtained by differentiating LL function with respect to q and α as

$$\begin{aligned} \frac{\partial l}{\partial q} &= -\frac{n}{1-q} + \frac{1}{q} \sum_{i=1}^n y_i + \sum_{i=1}^n \frac{\alpha q^{y_i} + \alpha y_i(1+q)q^{y_i-1}}{1-\alpha + \alpha(1+q)q^{y_i}}, \\ \frac{\partial l}{\partial \alpha} &= \sum_{i=1}^n \frac{(1+q)q^{y_i} - 1}{1-\alpha + \alpha(1+q)q^{y_i}}. \end{aligned}$$

The MLE ($\hat{\Theta}$) of Θ is obtained by solving the nonlinear system of equation $U(\Theta, \mathbf{y}) = 0$. Since the likelihood equations have no closed form solution, the estimator \hat{q} and $\hat{\alpha}$ of the parameters q and α can be obtained by maximizing LL function using global numerical maximization techniques using `optim` package of R. Further, the Fishers information matrix is given by

$$I_{\mathbf{y}}(q, \alpha) = \begin{pmatrix} -\mathbb{E}\left(\frac{\partial^2 l}{\partial q^2}\right) & -\mathbb{E}\left(\frac{\partial^2 l}{\partial q \partial \alpha}\right) \\ -\mathbb{E}\left(\frac{\partial^2 l}{\partial q \partial \alpha}\right) & -\mathbb{E}\left(\frac{\partial^2 l}{\partial \alpha^2}\right) \end{pmatrix} \approx \begin{pmatrix} -\frac{\partial^2 l}{\partial q^2} & -\frac{\partial^2 l}{\partial q \partial \alpha} \\ -\frac{\partial^2 l}{\partial q \partial \alpha} & -\frac{\partial^2 l}{\partial \alpha^2} \end{pmatrix}_{q=\hat{q}, \alpha=\hat{\alpha}}, \tag{5}$$

where \hat{q} and $\hat{\alpha}$ are the MLE's of q and α respectively. Moreover elements of $I_{\mathbf{y}}(q, \alpha)$ may be computed from

$$\begin{aligned} \frac{\partial^2 l}{\partial q^2} &= -\frac{n}{(1-q)^2} - \frac{1}{q^2} \sum_{i=1}^n y_i - \sum_{i=1}^n \left(\frac{\alpha(1+q)(y_i-1)y_i q^{y_i-2} + 2\alpha y_i q^{y_i-1}}{1-\alpha + \alpha(1+q)q^{y_i}} - \left(\frac{\alpha(1+q)y_i q^{y_i-1} + \alpha q^{y_i}}{1-\alpha + \alpha(1+q)q^{y_i}} \right)^2 \right), \\ \frac{\partial^2 l}{\partial q \partial \alpha} &= \sum_{i=1}^n \left(\frac{(1+q)y_i q^{y_i-1} + q^{y_i}}{1-\alpha + \alpha(1+q)q^{y_i}} - \frac{(\alpha(1+q)y_i q^{y_i-1} + \alpha q^{y_i})((1+q)q^{y_i} - 1)}{(1-\alpha + \alpha(1+q)q^{y_i})^2} \right), \\ \frac{\partial^2 l}{\partial \alpha^2} &= -\sum_{i=1}^n \left(\frac{((1+q)q^{y_i} - 1)^2}{1-\alpha + \alpha(1+q)q^{y_i}} \right). \end{aligned}$$

4.5. MLE through EM Algorithm

The Expected Maximization (EM) algorithm is an useful iterative procedure to compute ML estimators in the presence of missing data or assumed to have a missing values. The procedure follows with two steps called Expectation step (E-Step) and Maximization step (M-Step). The E-step concerns with the estimation of those data which are not observed whereas the M-step is a maximization step for more details one may refer Dempster *et al.* [5]. Moreover, theorem 3.1 of Redner and Walker [6] ensures the consistency and uniqueness of the estimates obtained by EM procedure.

Let the complete-data be constituted with observed set of values $\mathbf{y} = (y_1, \dots, y_n)$ and the hypothetical data set $\mathbf{x} = (x_1, \dots, x_n)$, where the observations y_i 's are distributed with rv X defined as

$$X = \begin{cases} 1 & \text{with probability } (1 + \alpha)/2 \\ 0 & \text{with probability } (1 - \alpha)/2 \end{cases}, \quad (6)$$

and rv Y be defined as

$$Y = XZ_{1:2} + (1 - X)Z_{2:2}, \quad (7)$$

where $Z_{1:2} \sim GD(q^2)$, $Z_{2:2} \sim EGD(q, 2)$ (see [7]) and $X_i \sim \text{Bernoulli}\left(\frac{1+\alpha}{2}\right)$.

Under the formulation, the **E-step** of an EM cycle requires the expectation of $(X|Y; \Theta^{(k)})$, where $\Theta^{(k)} = (q^{(k)}, \alpha^{(k)})$ is the current estimate of Θ (in the k^{th} iteration). Since the conditional distribution of X_i given Y_i is

$$(X_i|Y_i, \Theta^{(k)}) \sim \text{bernoulli}\left(\frac{1 + \alpha_i^{(k)}}{2}\right), \quad (8)$$

with

$$\frac{1 + \alpha_i^{(k)}}{2} = \frac{(1 + \alpha^{(k)}) (1 - (q^{(k)})^2) (q^{(k)})^{2y_i}}{(1 + \alpha^{(k)}) (1 - (q^{(k)})^2) (q^{(k)})^{2y_i} + (1 - \alpha^{(k)}) ((1 - q^{(k)}) (q^{(k)})^{y_i} (2 - (q^{(k)} + 1) (q^{(k)})^{y_i}))}, \quad (9)$$

where $\alpha^{(k)}$ is a set of known or estimated parameters at k^{th} step with known initial values. Thus, by the property of the Binomial distribution, the conditional mean is

$$\mathbb{E}(X_i|Y_i, \Theta^{(k)}) = \left(\frac{1 + \alpha_i^{(k)}}{2}\right) \quad \text{and} \quad \mathbb{V}(X_i|Y_i, \Theta^{(k)}) = \left(\frac{1 + \alpha_i^{(k)}}{2}\right) \left(\frac{1 - \alpha_i^{(k)}}{2}\right). \quad (10)$$

For **M-step**: The likelihood function of joint pmf of hypothetical complete-data (Y_i, X_i) , $i = 1, \dots, n$ is given as

$$L^*(\Theta; \mathbf{y}, \mathbf{x}) = \prod_{i=1}^n \left(\frac{1 + \alpha}{2}\right)^{x_i} ((1 - q^2) q^{2y_i})^{x_i} \cdot \prod_{i=1}^n \left(\frac{1 - \alpha}{2}\right)^{1-x_i} (((1 - q) (q)^{y_i} (2 - (1 + q) q^{y_i})))^{1-x_i}$$

and the corresponding complete LL function is given as

$$l_n^*(\Theta; \mathbf{x}, \mathbf{y}) = \log\left(\frac{1 + \alpha}{2}\right) \sum_{i=1}^n x_i + \log\left(\frac{1 - \alpha}{2}\right) \sum_{i=1}^n (1 - x_i) + \log(1 - q^2) \sum_{i=1}^n x_i + 2 \log q \sum_{i=1}^n x_i y_i + \sum_{i=1}^n (1 - x_i) (y_i \log q + \log(1 - q) + \log(2 - q^{y_i} (1 + q))). \quad (11)$$

The components of the score function $U_n^*(\Theta) = \left(\frac{\partial l_n^*}{\partial \alpha}, \frac{\partial l_n^*}{\partial q}\right)^\top$ are given by

$$\frac{\partial l_n^*}{\partial \alpha} = \frac{1}{1 + \alpha} \sum_{i=1}^n x_i - \frac{1}{1 - \alpha} \sum_{i=1}^n (1 - x_i), \quad (12)$$

$$\begin{aligned} \frac{\partial I_n^*}{\partial q} = & -\frac{2q}{1-q^2} \sum_{i=1}^n x_i + \frac{2}{q} \sum_{i=1}^n x_i y_i \\ & + \sum_{i=1}^n (1-x_i) \left(\frac{y_i}{q} - \frac{1}{1-q} - \frac{y_i q^{y_i-1} + (y_i+1) q^{y_i}}{2-q^{y_i}(1+q)} \right). \end{aligned} \tag{13}$$

The EM cycle will be completed with the M-step by using the MLE over Θ , that is, $U_n^*(\hat{\Theta}; \mathbf{y}, \mathbf{x}) = 0$ with the unobserved x_i s replaced by their conditional expectations given in (10). Hence we obtain the iterative procedure of the EM algorithm as

$$\begin{aligned} \hat{\alpha}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n \alpha_i^{(k)}, \\ \hat{q}^{(k+1)} &= \frac{\sum_{i=1}^n \left(\frac{1+\alpha_i^{(k)}}{2} \right) y_i}{\frac{2\hat{q}^{(k+1)}}{1-(\hat{q}^{(k+1)})^2} \sum_{i=1}^n \left(\frac{1+\alpha_i^{(k)}}{2} \right) - \sum_{i=1}^n \left(\frac{1-\alpha_i^{(k)}}{2} \right) \left(\frac{y_i}{\hat{q}^{(k+1)}} - \frac{1}{1-\hat{q}^{(k+1)}} - \frac{y_i (\hat{q}^{(k+1)})^{y_i-1} + (y_i+1) (\hat{q}^{(k+1)})^{y_i}}{2-(\hat{q}^{(k+1)})^{y_i}(1+\hat{q}^{(k+1)})} \right)}, \end{aligned}$$

where $\hat{q}^{(k+1)}$ should be determined numerically.

4.5.1. Standard errors of estimates obtained from EM algorithm

In this section, we obtain the standard errors (se) of the estimators from the EM algorithm using result of Louis [8]. Let $\mathbf{z} = (\mathbf{y}, \mathbf{x})$, then the elements of 2×2 observed information matrix $I_c(\Theta, \mathbf{z}) = \left[\frac{\partial}{\partial \Theta} U_c(\Theta; \mathbf{z}) \right]$ are given by

$$\begin{aligned} \frac{\partial^2 I_n^*}{\partial \alpha^2} &= -\frac{1}{(1+\alpha)^2} \sum_{i=1}^n x_i - \frac{1}{(1-\alpha)^2} \sum_{i=1}^n (1-x_i), \\ \frac{\partial^2 I_n^*}{\partial \alpha \partial q} &= \frac{\partial^2 I_n^*}{\partial q \partial \alpha} = 0, \\ \frac{\partial^2 I_n^*}{\partial q^2} &= -\frac{2(1+q^2)}{(1-q^2)^2} \sum_{i=1}^n x_i - \frac{2}{q^2} \sum_{i=1}^n x_i y_i - \sum_{i=1}^n (1-x_i) \left(\frac{y_i}{q^2} + \frac{q^{2y_i-2} (qy_i + q + y_i)^2}{(2-(q+1)q^{y_i})^2} \right. \\ &\quad \left. + \frac{(y_i-1)y_i q^{y_i-2} + y_i(y_i+1)q^{y_i-1}}{2-(q+1)q^{y_i}} + \frac{1}{(1-q)^2} \right). \end{aligned}$$

Taking the conditional expectation of $I_c(\Theta; \mathbf{z})$ given \mathbf{x} , we obtain the 2×2 matrix

$$l_c(\Theta; \mathbf{y}) = -\mathbb{E}(I_c(\Theta; \mathbf{z}) | \mathbf{y}) = (d_{ij}), \tag{14}$$

where

$$\begin{aligned} d_{11} &= \frac{1}{(1+\alpha)^2} \sum_{i=1}^n \mathbb{E}(X_i | \mathbf{y}) + \frac{1}{(1-\alpha)^2} \sum_{i=1}^n (1 - \mathbb{E}(X_i | \mathbf{y})), \\ d_{12} &= d_{21} = 0, \\ d_{22} &= \frac{2(1+q^2)}{(1-q^2)^2} \sum_{i=1}^n \mathbb{E}(X_i | \mathbf{y}) + \frac{2}{q^2} \sum_{i=1}^n \mathbb{E}(X_i | \mathbf{y}) y_i \\ &\quad + \sum_{i=1}^n (1 - \mathbb{E}(X_i | \mathbf{y})) \left(\frac{y_i}{q^2} + \frac{q^{2y_i-2} (qy_i + q + y_i)^2}{(2-(q+1)q^{y_i})^2} \right. \\ &\quad \left. + \frac{(y_i-1)y_i q^{y_i-2} + y_i(y_i+1)q^{y_i-1}}{2-(q+1)q^{y_i}} + \frac{1}{(1-q)^2} \right). \end{aligned}$$

whereas computation of

$$l_m(\Theta; \mathbf{y}) = \mathbb{V}(U_c(x; \theta) | \mathbf{y}) = m_{ij}, \tag{15}$$

involve the following terms

$$m_{11} = \left(\frac{1}{1+\alpha} + \frac{1}{1-\alpha} \right)^2 \sum_{i=1}^n \mathbb{V}(X_i | \mathbf{y}),$$

$$m_{12} = m_{21} = \sum_{i=1}^n \left(\frac{1}{1+\alpha} + \frac{1}{1-\alpha} \right) \left(\frac{y_i}{q} - \frac{2q}{1-q^2} + \frac{1}{1-q} + \frac{y_i q^{y_i-1} + (y_i+1) q^{y_i}}{2 - q^{y_i} (1+q)} \right) \mathbb{V}(X_i | \mathbf{y}),$$

$$m_{22} = \sum_{i=1}^n \left(\frac{y_i}{q} - \frac{2q}{1-q^2} + \frac{1}{1-q} + \frac{y_i q^{y_i-1} + (y_i+1) q^{y_i}}{2 - q^{y_i} (1+q)} \right)^2 \mathbb{V}(X_i | \mathbf{y}).$$

Finally, the observed information matrix (I) can be computed as

$$I(\hat{\Theta}; \mathbf{y}) = l_c(\hat{\Theta}; \mathbf{y}) - l_m(\hat{\Theta}; \mathbf{y}),$$

and $I(\hat{\Theta}; \mathbf{y})$ can be inverted to obtain an estimate of the covariance matrix of the incomplete-data problem. The square roots of the diagonal elements represent the estimates of the standard errors of the parameters.

4.6. Simulation Study to Evaluate EM Algorithm

Here we study the behavior of ML estimators obtained by direct numerical optimization and also through EM algorithm for different finite sample sizes and for different $TGD(q, \alpha)$. Observations from $TGD(q, \alpha)$ are generated using the quantile function provided in Chakraborty and Bhati [1] (see result 4 of Table 1). In the next two subsections, first we investigate the performance of ML estimators $(\hat{q}, \hat{\alpha})$ for various combinations of parameters (q, α) in subsection (3.6.1) and then evaluate the performance with respect to varying sample size for fixed parameter values in subsection (3.6.2).

4.6.1. Performance of estimators for different parametric values

A simulation study consisting of following steps is carried out for each triplet (q, α, n) , considering $q = 0.25, 0.5, 0.75$, $\alpha = -0.70, -0.30, 0.30, 0.70$ and $n = 25, 50, 75, 100$.

1. Choose the value (q_0, α_0) for the corresponding elements of the parameter vector $\Theta = (q, \alpha)$, to specify the $TGD(q, \alpha)$;
2. Choose sample size n ;
3. Generate N independent samples of size n from $TGD(q, \alpha)$;
4. Compute the ML and EM estimate $\hat{\Theta}_n$ of Θ for each of the N samples;
5. Compute the average bias, standard error (SE) of the estimate.

In our experiment we have considered the number of replication $N = 1000$. It can be observed from Tables 1 and 2 that as the sample size increases both average bias and average se decreases.

4.6.2. Performance of estimators for different sample size

In this subsection, we assess the performance of ML estimators of (q, α) as sample size n , increases by considering $n = 25, 26, \dots, 200$, for $q = 0.25$ and $\alpha = -0.5$. For each n , we generate one thousand samples of size n and obtain MLEs and their standard error. For each repetition we compute average bias and average squared error.

Figures 2 and 3 show behavior of average bias and average standard error of parameter q and α , for fixed $q = 0.25$ and $\alpha = -0.5$, as one varies sample size n . The horizontal dotted lines from Figure 2 corresponds to zero value and it is clear in Figure 2 that the biases approach to zero with increasing n . Further, in Figure 3, average standard errors for both parameters (q and α) decrease with increase in n . Similar observations were also noted for other parametric values.

Based on our findings it is clear that EM algorithm produces better ML estimators with smaller average bias as compared to the regular ML estimators while w.r.t. standard error there is not much to choose between the two procedures.

Table 1 | Average Bias and Average SE computed by method of maximum likelihood and EM Algorithm method.

Parameters		n	MLE				EM Algorithm			
			bias($\hat{\alpha}$)	bias(\hat{q})	se($\hat{\alpha}$)	se(\hat{q})	bias($\hat{\alpha}$)	bias(\hat{q})	se($\hat{\alpha}$)	se(\hat{q})
q = 0.25	$\alpha = -0.75$	25	-0.5566	0.0099	1.5154	0.1144	-0.0132	0.0232	0.9739	0.1147
		50	-0.2675	0.0101	0.9151	0.0866	-0.0081	0.0122	0.7215	0.0835
		75	-0.1733	0.0050	0.6880	0.0694	-0.0049	0.0073	0.5881	0.0677
		100	-0.1327	0.0053	0.5780	0.0600	-0.0035	0.0052	0.5137	0.0589
q = 0.5	$\alpha = -0.75$	25	-0.1348	-0.0149	0.5644	0.0859	-0.0031	-0.0058	0.5664	0.0854
		50	-0.0077	-0.0001	0.3960	0.0619	0.0012	-0.0029	0.3888	0.0601
		75	-0.0196	-0.0012	0.3197	0.0498	-0.0006	-0.0011	0.3155	0.0489
		100	-0.0113	-0.0026	0.2765	0.0432	0.0012	-0.0028	0.2730	0.0424
q = 0.75	$\alpha = -0.75$	25	-0.0411	-0.0003	0.4012	0.0480	0.0060	-0.0035	0.4190	0.0476
		50	-0.0085	-0.0026	0.2766	0.0333	-0.0002	-0.0021	0.2964	0.0333
		75	-0.0011	-0.0018	0.2242	0.0268	0.0012	-0.0020	0.2337	0.0268
		100	-0.0008	-0.0014	0.1909	0.0227	-0.0009	-0.0007	0.1990	0.0229
q = 0.25	$\alpha = -0.30$	25	-0.3455	0.0269	1.2484	0.1361	-0.0340	0.0260	1.0043	0.1426
		50	-0.3055	0.0095	0.9006	0.1016	-0.0240	0.0165	0.7203	0.1018
		75	-0.0391	0.0290	0.6466	0.0901	-0.0125	0.0109	0.6045	0.0848
		100	-0.0997	0.0123	0.5770	0.0756	-0.0090	0.0069	0.5269	0.0736
q = 0.5	$\alpha = -0.30$	25	-0.0310	0.0045	0.6288	0.1097	-0.0037	0.0000	0.6840	0.1153
		50	-0.0249	0.0011	0.4672	0.0803	-0.0031	-0.0009	0.4818	0.0818
		75	-0.0253	-0.0002	0.3880	0.0668	-0.0029	-0.0008	0.3908	0.0664
		100	-0.0258	-0.0008	0.3375	0.0580	-0.0036	-0.0004	0.3333	0.0568
q = 0.75	$\alpha = -0.30$	25	-0.0503	0.0000	0.5182	0.0625	-0.0010	-0.0046	0.6044	0.0689
		50	-0.0432	0.0010	0.3850	0.0459	-0.0069	-0.0030	0.4157	0.0482
		75	-0.0020	0.0003	0.3141	0.0369	-0.0038	0.0000	0.3324	0.0381
		100	-0.0009	0.0000	0.2838	0.0330	-0.0026	-0.0005	0.2899	0.0332

Note: EM = Expected Maximization; MLE = maximum likelihood estimator.

Table 2 | Average Bias and Average SE computed by method of maximum likelihood and EM Algorithm method.

Parameters		n	MLE				EM Algorithm			
			bias($\hat{\alpha}$)	bias(\hat{q})	se($\hat{\alpha}$)	se(\hat{q})	bias($\hat{\alpha}$)	bias(\hat{q})	se($\hat{\alpha}$)	se(\hat{q})
q = 0.25	$\alpha = 0.30$	25	-0.4174	0.0254	1.1108	0.1524	-0.1138	-0.0206	0.7619	0.1547
		50	-0.2518	0.0178	0.8702	0.1281	-0.0667	-0.0158	0.8095	0.1632
		75	-0.1338	0.0193	0.6331	0.1110	-0.0481	-0.0144	0.6889	0.1386
		100	-0.0878	0.0215	0.5479	0.1013	-0.0367	-0.0032	0.6740	0.1353
q = 0.50	$\alpha = 0.30$	25	-0.2343	0.0226	0.5962	0.1328	-0.0404	-0.0267	0.7990	0.1700
		50	-0.1440	0.0184	0.4884	0.1085	-0.0335	-0.0354	0.6296	0.1349
		75	-0.0611	0.0142	0.4132	0.0926	-0.0319	-0.0336	0.5801	0.1237
		100	-0.0586	0.0125	0.3970	0.0886	-0.0213	-0.0210	0.5013	0.1072
q = 0.75	$\alpha = 0.30$	25	-0.0594	0.0127	0.5713	0.0829	-0.0143	-0.0326	0.7882	0.1101
		50	-0.0316	0.0097	0.4540	0.0652	-0.0173	-0.0508	0.6689	0.0923
		75	-0.0250	0.0079	0.3969	0.0568	-0.0177	-0.0607	0.5449	0.0759
		100	-0.0081	0.0050	0.3729	0.0522	-0.0107	-0.0224	0.5029	0.0691
q = 0.25	$\alpha = 0.75$	25	-0.0975	0.0038	0.0240	0.0042	-0.0234	-0.0305	0.6862	0.1166
		50	-0.0696	0.0138	0.0255	0.0027	-0.0189	-0.0220	0.5239	0.0423
		75	-0.0995	0.0046	0.0751	0.0038	-0.0125	-0.0112	0.4443	0.0259
		100	-0.0358	0.0070	0.0338	0.0027	-0.0101	-0.0071	0.4012	0.0125
q = 0.5	$\alpha = 0.75$	25	-0.1250	0.0288	0.5170	0.1474	-0.0351	-0.0112	0.5214	0.1627
		50	-0.1162	0.0248	0.4238	0.1186	-0.0158	-0.0131	0.4862	0.1456
		75	-0.0641	0.0140	0.3485	0.1000	-0.0093	-0.0583	0.3675	0.1088
		100	-0.0493	0.0125	0.3422	0.0974	-0.0037	-0.1109	0.6810	0.1963
q = 0.75	$\alpha = 0.75$	25	-0.1542	0.0350	0.5112	0.0966	-0.0191	-0.0176	0.5221	0.1048
		50	-0.1114	0.0178	0.4014	0.0727	-0.0350	-0.0253	0.4722	0.0858
		75	-0.0786	0.0100	0.3595	0.0638	-0.1072	-0.0986	0.3782	0.0676
		100	-0.0455	0.0100	0.3139	0.0566	-0.1168	-0.1178	0.3662	0.0647

Note: EM = Expected Maximization; MLE = maximum likelihood estimator.

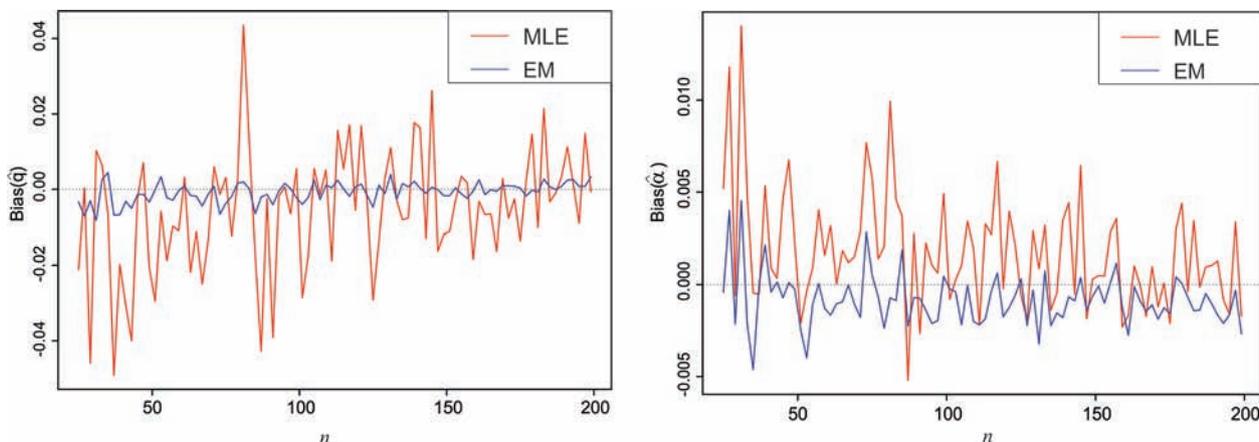


Figure 2 | Bias plot of estimated value of parameter q and α for different sample sizes.

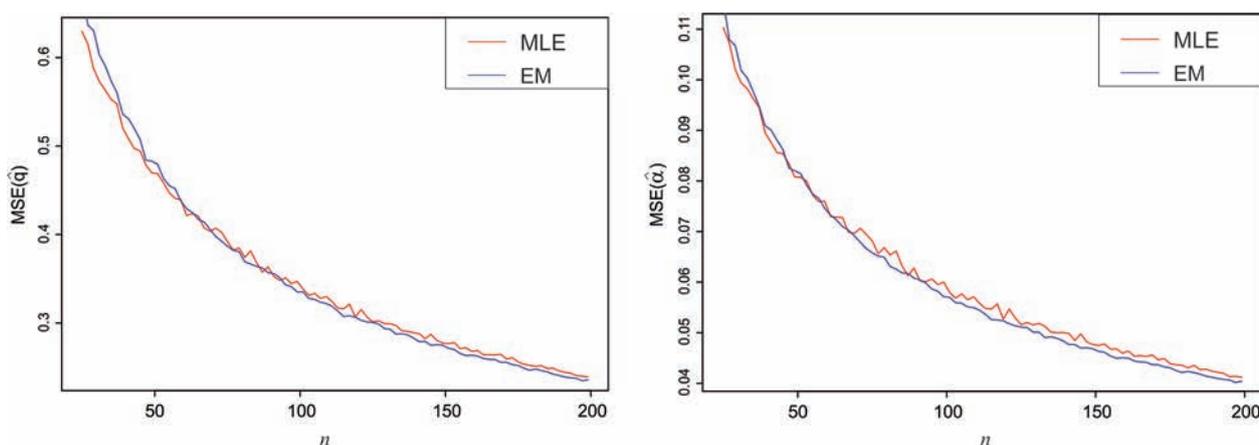


Figure 3 | MSE plot of estimated value of parameter q and α for different sample sizes.

5. TESTS OF HYPOTHESIS

The $TGD(q, \alpha)$ distribution with parameter vector $\Theta = (q, \alpha)^T$ reduces to the Geometric distribution with parameter q when $\alpha = 0$. This additional parameter α controls the proportion of zeros of the distribution relative to geometric distribution and also the tail length. Therefore it is of interest to develop test procedure for detecting departure of α from 0. In this section we develop the LRT, the Rao’s score test and the Wald’s test for testing the null hypothesis $H_0 : \alpha = 0$ against the alternative hypothesis $H_1 : \alpha \neq 0$ and numerically study the statistical power of these tests through extensive simulation.

5.1. LRT, Rao’s Score Test and Wald’s Test

The LRT is based on the difference between the maximum of the likelihood under null and the alternative hypotheses. The LRT statistics is given by $-2 \log \left(\frac{L(\hat{\Theta}^*; y)}{L(\hat{\Theta}; y)} \right)$ where $\hat{\Theta}^*$ and $\hat{\Theta}$ are the MLE obtained under the null and alternative hypotheses respectively. The LRT is generally employed to test the significance of the additional parameter which is included to extend a base model.

The Rao’s Score test [9] is based on the score vector defined as the first derivative of the LL function w.r.t. the parameters. Rao’s score test statistic $UI^{-1}U^T$, where U is the score vector and I is the information matrix derived under the null hypothesis. The score vector and the information matrix, obtained by evaluating the derivatives of the LL function, $\log L$ is provided in Section 4.4. Note that the scores actually are the slopes of the likelihood functions.

The Wald’s test statistics [10] is based on the difference between the maximum of the likelihood estimate value of the parameter under alternative hypothesis and the value of specified null hypothesis. The Wald’s test statistic is given in our case by $(\hat{\alpha} - \alpha_0) I_{[22]}^{-1} (\hat{\alpha} - \alpha_0)^T$, where $I_{[22]}^{-1}$ is the (2, 2)th element of the inverse of the information matrix I , and $\hat{\alpha}$ is the MLE of α both under alternative hypotheses, whereas α_0 is the value of α as per H_0 . Note that $I_{[22]}^{-1}$ is an estimate of the variance of α . Therefore in the present case our Wald’s statistic reduces to $(\hat{\alpha})^2 \mathbb{V}(\hat{\alpha})$.

All the test statistics follow asymptotically chi-square distribution with “ k ” degrees of freedom (df), where “ k ” is the number of parameters specified by the null hypothesis. In the present case the df is just “1”. For well behaved likelihood function all these tests are based on measuring the discrepancy between null and the alternative hypotheses.

5.2. Statistical Power Analysis

Here we present a simulation based study of the statistical power of LR test, Rao’s Score test and Wald’s test considering 5% level of significance. Since the tests are asymptotic in nature we have considered four different sample sizes, namely $n = 100, 300, 500$ and 1000 . We have generated 1000 replications for each sample size n . The power of these tests are estimated by proportion of rejection in these 1000 replications. The effect size (ES) is a measure of departure from the null hypothesis which in the present case is given by $\alpha - 0 = \alpha$ is fixed at $-0.7, -0.5, -0.3, -0.1, 0.1, 0.2, 0.5, 0.7$ for our experiments.

The power of the these test for different sample size, ES and for different parametric value q are presented in Tables 3 and 4. Figures 4 and 5 reveal, as expected, that the power of the test increases with the sample size n and ES. Further, for positive ES, all the tests show increase in power with the increase in either or both ES and sample size, Moreover, for negative ES, power increase in a much faster pace. Power for score test is more than LR test for negative ES where as it is other way for positive ES. For positive ES the power of the tests gets closer with increase in sample size. From the over all observation it is clear that the Wald’s test is more reliable than both LR test and Score tests.

Table 3 | Power of the LR, Score and Walds Test for different sample sizes n , effect size and parametric value q .

$q = 0.10$												
n	100			300			500			1000		
α	LR	Score	Wald									
-0.7	0.028	0.051	0.042	0.117	0.066	0.017	0.252	0.133	0.01	0.476	0.351	0.087
-0.5	0.02	0.046	0.041	0.052	0.039	0.029	0.121	0.066	0.024	0.237	0.134	0.015
-0.3	0.022	0.081	0.056	0.024	0.032	0.037	0.059	0.035	0.034	0.101	0.045	0.019
-0.1	0.013	0.11	0.066	0.02	0.069	0.061	0.036	0.045	0.046	0.059	0.04	0.032
0.1	0.015	0.173	0.074	0.021	0.123	0.109	0.042	0.112	0.099	0.058	0.105	0.071
0.3	0.009	0.239	0.091	0.026	0.209	0.143	0.038	0.212	0.161	0.079	0.237	0.134
0.5	0.005	0.28	0.112	0.041	0.351	0.217	0.069	0.394	0.266	0.142	0.431	0.288
0.7	0.033	0.262	0.132	0.025	0.452	0.236	0.074	0.523	0.336	0.197	0.602	0.436
$q = 0.30$												
n	100			300			500			1000		
α	LR	Score	Wald									
-0.7	0.305	0.565	0.127	0.742	0.851	0.741	0.922	0.963	0.927	0.998	0.999	0.999
-0.5	0.137	0.303	0.047	0.412	0.537	0.389	0.619	0.707	0.620	0.917	0.942	0.924
-0.3	0.074	0.177	0.047	0.147	0.219	0.129	0.207	0.274	0.200	0.457	0.519	0.462
-0.1	0.049	0.098	0.064	0.059	0.076	0.065	0.065	0.076	0.053	0.089	0.109	0.085
0.1	0.041	0.072	0.086	0.052	0.055	0.101	0.056	0.054	0.060	0.080	0.071	0.051
0.3	0.034	0.090	0.129	0.082	0.101	0.151	0.153	0.156	0.180	0.296	0.292	0.202
0.5	0.043	0.153	0.172	0.139	0.213	0.271	0.289	0.336	0.347	0.546	0.575	0.455
0.7	0.276	0.181	0.468	0.265	0.300	0.555	0.367	0.424	0.627	0.634	0.642	0.725
$q = 0.45$												
n	100			300			500			1000		
α	LR	Score	Wald									
-0.7	0.470	0.787	0.563	0.933	0.982	0.956	0.989	0.997	0.993	1.000	1.000	1.000
-0.5	0.241	0.540	0.310	0.611	0.792	0.675	0.835	0.909	0.873	0.993	0.996	0.994
-0.3	0.089	0.279	0.125	0.223	0.368	0.280	0.325	0.465	0.391	0.641	0.729	0.699
-0.1	0.058	0.157	0.089	0.076	0.128	0.105	0.071	0.106	0.085	0.090	0.137	0.122
0.1	0.035	0.083	0.078	0.060	0.062	0.096	0.059	0.055	0.071	0.097	0.073	0.051
0.3	0.033	0.062	0.117	0.117	0.083	0.171	0.210	0.163	0.193	0.396	0.313	0.233
0.5	0.055	0.106	0.199	0.224	0.200	0.316	0.417	0.351	0.427	0.700	0.645	0.539
0.7	0.268	0.121	0.468	0.347	0.227	0.639	0.497	0.377	0.711	0.763	0.685	0.805

Note: LR = likelihood ratio.

Table 4 | Power of the LR, Score and Walds Test for different sample sizes n , effect size and parametric value q .

$q = 0.6$												
n	100			300			500			1000		
α	LR	Score	Wald									
-0.7	0.628	0.888	0.760	0.985	0.997	0.996	1.000	1.000	1.000	1.000	1.000	1.000
-0.5	0.281	0.630	0.434	0.745	0.875	0.825	0.920	0.962	0.947	0.997	0.998	0.998
-0.3	0.120	0.351	0.213	0.273	0.457	0.364	0.453	0.627	0.550	0.748	0.830	0.802
-0.1	0.045	0.178	0.113	0.070	0.139	0.110	0.081	0.125	0.108	0.109	0.180	0.150
0.1	0.042	0.103	0.108	0.046	0.054	0.097	0.072	0.046	0.078	0.113	0.080	0.057
0.3	0.043	0.077	0.141	0.143	0.082	0.195	0.267	0.165	0.208	0.450	0.358	0.257
0.5	0.064	0.089	0.202	0.252	0.172	0.350	0.481	0.336	0.415	0.784	0.689	0.583
0.7	0.265	0.083	0.485	0.392	0.188	0.667	0.563	0.387	0.741	0.817	0.697	0.863

$q = 75$												
n	100			300			500			1000		
α	LR	Score	Wald									
-0.7	0.698	0.941	0.852	0.994	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000
-0.5	0.336	0.686	0.532	0.798	0.921	0.868	0.956	0.987	0.979	0.998	0.999	0.999
-0.3	0.142	0.374	0.245	0.297	0.512	0.418	0.488	0.669	0.600	0.810	0.877	0.860
-0.1	0.045	0.188	0.131	0.057	0.150	0.119	0.077	0.153	0.130	0.090	0.167	0.141
0.1	0.045	0.112	0.115	0.060	0.056	0.110	0.095	0.071	0.104	0.092	0.057	0.045
0.3	0.044	0.072	0.139	0.145	0.076	0.179	0.299	0.182	0.236	0.470	0.356	0.245
0.5	0.091	0.078	0.242	0.285	0.164	0.355	0.491	0.321	0.427	0.789	0.683	0.586
0.7	0.316	0.089	0.525	0.412	0.176	0.657	0.585	0.369	0.764	0.845	0.698	0.860

Note: LR = likelihood ratio.

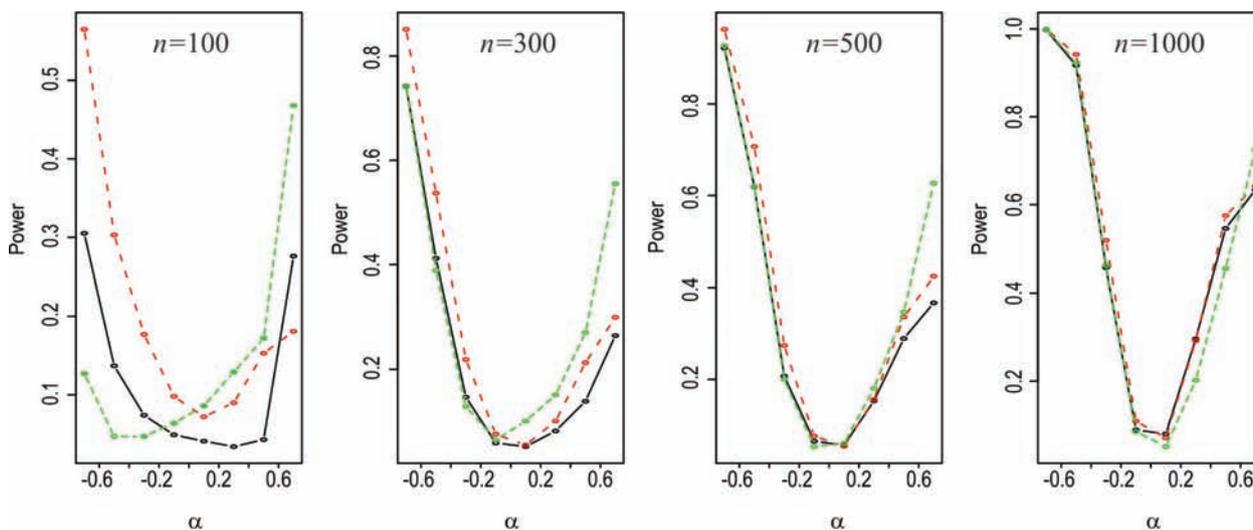


Figure 4 | Power curve of likelihood ratio (LR) test (black), Score test (Red) and Wald's test (Green) for different n and $q = 0.3$.

6. DATA ANALYSIS

For the purpose of illustration, in this section, we consider following two data sets with details as follows:

1. Number of Fires in Greece (NTG)

The data comprise of numbers of fires in district forest of Greece from period 1 July 1998 to 31 August 1998. The observed sample values of size 123 for these data are the following (frequency in parentheses and none when it is equal to one): 0(16), 1(13), 2(14), 3(9), 4(11), 5(13), 6(8), 7(4), 8(9), 9(6), 10(3), 11(4), 12(6), 15(4), 16, 20, 43. The data were previously studied by Bakouch *et al.* [11] and Karlis and Xekalaki [12].

2. Insurance Claim Count (ICC)

An insurance count data from Belgium in year 1993 is considered [13] and the data is as follows: 0(57178), 1(5617), 2(446), 3(50), 4(8).

The null hypothesis $H_0 : \alpha = 0$ against $H_1 : \alpha \neq 0$ are examined utilizing the LR, Rao’s Score and Wald’s test, and the results along with the descriptive statistics and with respective p -values are presented in Table 5. Moreover, p -value (less than 0.05) for Rao’s Score and Wald’s test reject the null hypothesis at 5% significance level. The suitability of the proposed $TGD(q, \alpha)$ model with other competitive distributions namely weighted geometric distribution ($WGD(q, \alpha)$) (Bhati and Joshi [14]), Negative Binomial ($NB(r, p)$), Poisson-weighted exponential distribution ($PWED(\alpha, \theta)$) (see [15]) and $GD(q)$ is carried out and the LL, Akaiki Information Criteria (AIC), Bayesian Information Criteria (BIC) value are computed for five models for both the datasets. Further the Kolmogorov-Smirnov goodness of fit value with bootstrap p -value is also computed. The results in Table 6 such as maximum LL, minimum AIC, BIC and KS value for $TGD(q, \alpha)$ reveals that the $TGD(q, \alpha)$ is the best fitted model and could be consider as competitive model for the datasets considered.

7. CONCLUSION

The current paper investigates some additional property of the $TGD(q, \alpha)$ distribution with emphasis on the simulation study of the behaviors of the parameter estimation and also power of tests of hypothesis to check statistical significance of the additional parameter. In the

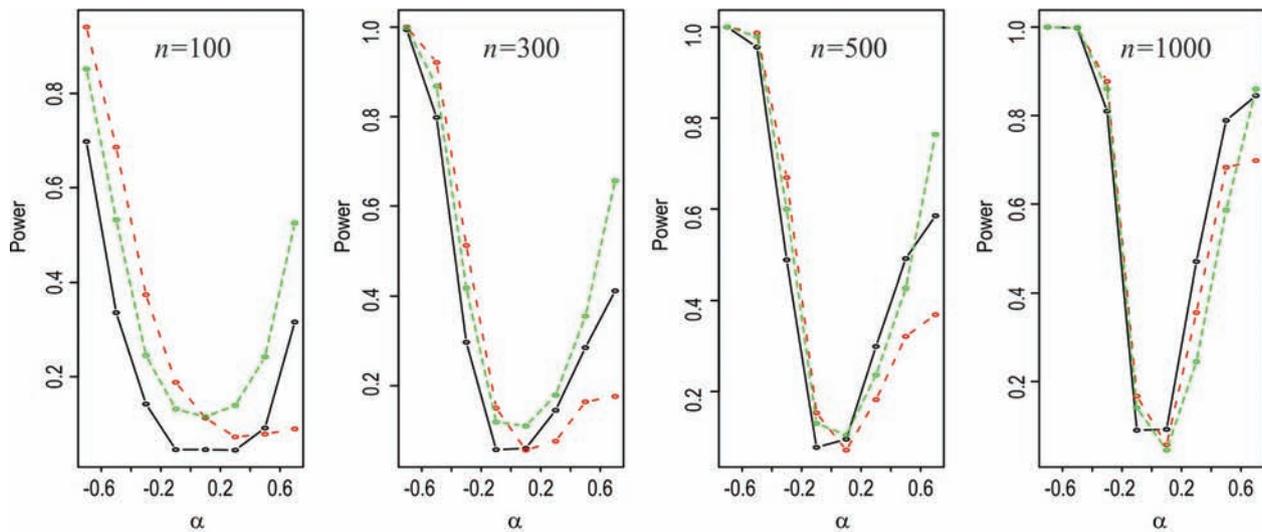


Figure 5 | Power of likelihood ratio (LR) test (black), Score test (Red) and Wald’s test (Green) for different n and $q = 0.75$.

Table 5 | Descriptive and test statistic with p -value in parentheses for both the datasets.

Data Set	Mean	Variance	LR test	Score Test	Wald’s Test
NTG	5.398	30.045	3.568 (0.059)	41.018 (0.000)	5.423 (0.020)
ICC	0.106	0.115	9.178 (0.002)	8.268 (0.004)	8.118 (0.004)

Notes: ICC = insurance claim count; LR = likelihood ratio; NTG = number of fires in Greece.

Table 6 | Comparative study of data fitting.

Dataset ↓	Model →	$WGD(q, \alpha)$	$NB(r, p)$	$PWED(\alpha, \theta)$	$TGD(q, \alpha)$	$GD(q)$
NTG	Estimates	(0.802, 1.804)	(1.496, 0.223)	(2.470, 0.247)	(0.802, -0.531)	0.839
	LL	-335.552	-334.529	-335.552	-334.415	-337.165
	AIC	675.104	673.058	675.104	672.830	676.330
	BIC	680.728	678.682	680.728	678.454	679.142
	KS	0.547	0.449	0.525	0.431	0.906
	p -value	0.767	0.882	0.847	0.899	0.432
ICC	Estimates	(0.0845, 0.749)	(1.279, 0.924)	(5.870, 10.837)	(0.085, -0.157)	0.095
	LL	-22063.8	-22064.3	-22063.8	-22063.6	-22068.2
	AIC	44131.6	44132.6	44131.6	44131.2	44138.4
	BIC	44149.7	44150.7	44149.7	44149.3	44147.4
	KS	0.091	0.102	0.091	0.074	0.427
	p -value	0.878	0.862	0.882	0.918	0.525

Notes: AIC = Akaiki Information Criteria; BIC = Bayesian Information Criteria; ICC = insurance claim count; LL = log likelihood; NTG = number of fires in Greece.

parameter estimation we have presented different methods including the EM algorithm implementation of the MLE. A comparative simulation based evaluation of the EM algorithm based MLE against the usual MLE has revealed the superiority of the former in terms of the bias and mean squared errors. We have also presented data modeling examples to showcase the advantage of the $TGD(q, \alpha)$ over some of the existing distributions from literature. As such it is envisaged that the present contribution will be useful for discrete data analysts.

CONFLICT OF INTEREST

The authors declare they have no conflicts of interest.

AUTHORS' CONTRIBUTIONS

SC contributes in Section 1–3,5,7 whereas DB contributes in Section 4–7.

Funding Statement

This research is not funded by any agency.

ACKNOWLEDGMENT

The authors acknowledge with thanks the comments of two anonymous referees which helped immensely in improving the presentation of the paper.

REFERENCES

1. S. Chakraborty, D. Bhati, *Stat. Oper. Res. Trans.* 40 (2016), 153–176.
2. W. Shaw, I. Buckley, *The Alchemy of Probability Distributions: Beyond Gram-Charlier Expansions and a Skew-Kurtotic-Normal Distribution from a Rank Transmutation Map*, Research Report, 2007. arXiv:0901.0434v1 [q-fin.ST].
3. M. Shaked, J. Shanthikumar, *Stochastic Orders and Their Applications*. Probability and Mathematical Statistics, Academic Press, Boston, 1994.
4. M.S.A. Khan, A. Khalique, A.M. Aboummoh, *IEEE Trans. Reliab.* 38 (1989), 348–350.
5. A.P. Dempster, N.M. Laird, D.B. Rubin, *J. R. Stat. Soc. Ser. B.* 39 (1977), 1–38.
6. R.A. Redner, H.R. Walker, *SIAM Rev.* 26 (1984), 195–239.
7. S. Chakraborty, R.D. Gupta, *Exponentiated geometric distribution: another generalization of geometric distribution*, *Commun. Stat. Theory Methods.* 44 (2015), 1143–1157.
8. T. Louis, *J. R. Stat. Soc. Ser. B Methodol.* 44 (1982), 226–233.
9. C.R. Rao, *Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation*, *Math. Proc. Camb. Philos. Soc.* 44 (1948), 50–57.
10. A. Wald, *Trans. Am. Math. Soc.* 54 (1943), 426–482.
11. H.S. Bakouch, M.A. Jazi, S. Nadarajah, *Statistics.* 48 (2014), 200–240.
12. D. Karlis, E. Xekalaki, in: E.A. Lipitakis (Ed.), *Proceedings of the Fifth Hellenic-European Conference on Computer Mathematics and Its Applications*, 2001, pp. 872–877.
13. M. Denuit, *Astin Bull.* 27 (1997), 229–242.
14. D. Bhati, S. Joshi, *Commun. Stat. Theor. Methods.* 47 (2018), 1510–1527.
15. H. Zamani, N. Ismail, P. Faroughi, *J. Math. Stat.* 10 (2014), 148–154.